

Experiences with Data Parallel Frameworks

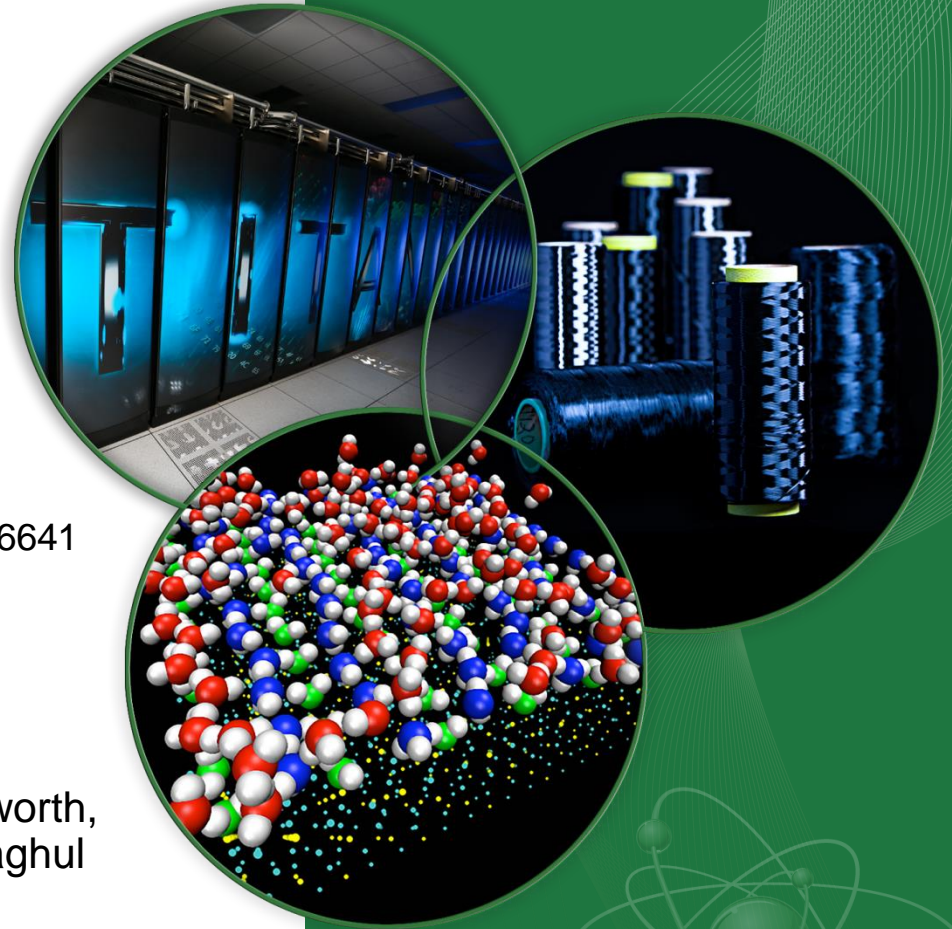
Sreenivas Rangan Sukumar, PhD

Email: sukumarsr@ornl.gov

Phone: 865-241-6641

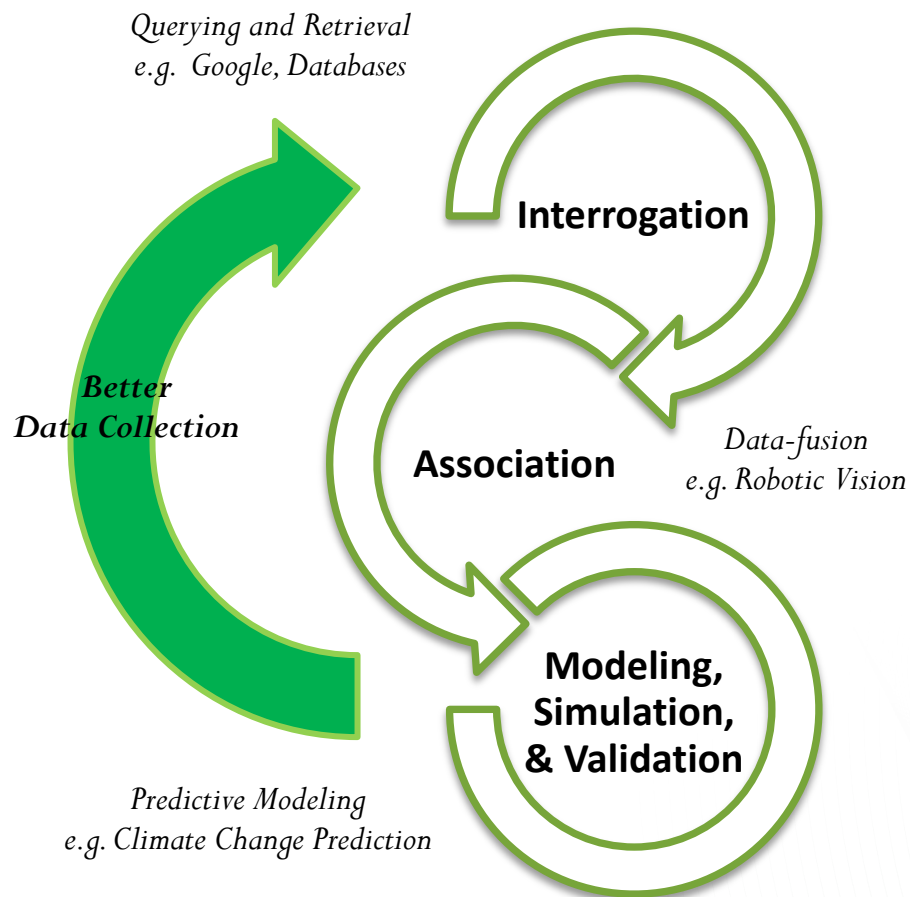
Team:

Seung-Hwan Lim, Chris Symons, Keela Ainsworth,
James Horey, Tyler Brown, Edmon Begoli, Raghul
Gunasekaran, Mallikarjun Shankar , Galen
Shipman and Jack Wells



Recap from yesterday....

The Lifecycle of Data-Intensive Discovery

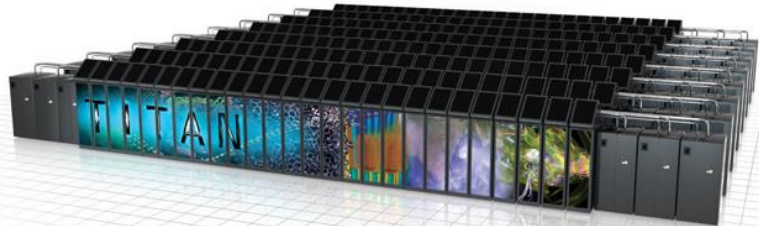


Can we scale up in all three aspects of discovery?

So what are we doing at ORNL ?

Understanding Data Parallel Frameworks for Data-intensive Computing

Self-taught learning on cores and GPUs ?



Multi-task learning at scale ?



Can cloud catch it all ?



Graph computing on shared-memory platforms ?

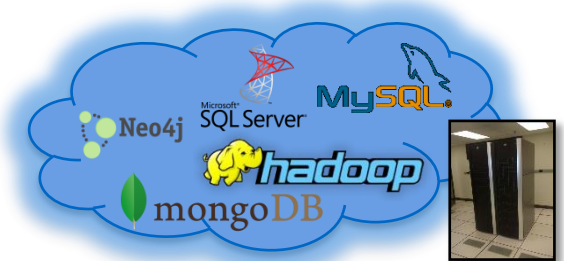


Running Hadoop on Eos and Rhea ?

Food-for-thought for the Exascale “Data-Analysis” Supercomputer

Architectures for Scalable Analytics

OLCF Compute Resources



	Titan	Apollo	Cloud
Discovery Approach	Modeling and Simulation	Association	Querying, Prediction
Architecture	Shared-nothing	Shared-memory	Shared-storage
Scalability	Compute (# of cores)	Horizontal (# of datasets)	Vertical (# of rows)
Algebra	Linear	Relational	Set-theoretic
Challenge (Pros)	Resolution	Heterogeneity	Cost
Challenge (Cons)	Dimensionality	Custom Solution	Flexibility
Leadership	#2 in the world (2013)	1 of 15 installs (2013)	--
User-interface	OpenMP, MPI, CUDA	SPARQL	SQL

Three Analytical-Support Toolkits

- Still in R&D stage (Use-cases and collaborations are very welcome)
 - EAGLE : The exploratory discovery module on ORNL's Apollo (YarcData Urika)
 - SpotHadoop: Hybridizing HPC and Big Data community (Tested on: Smoky, Rhea, EOS Troubleshooting: TITAN)
 - iLearn: Machine Learning using heterogeneous computing (Implementing self-learning methods such as deep-and multi-task learning on TITAN)

Three Analytical-Support Toolkits

- Still in R&D stage (Use-cases and ideas very welcome)
 - EAGLE : The exploratory discovery module on ORNL's Apollo (YarcData Urika)
 - SpotHadoop: Hybridizing HPC and Big Data community (Tested on: Smoky, Rhea, EOS Troubleshooting: TITAN)
 - iLearn: Machine Learning using heterogeneous computing (Implementing self-learning methods such as deep-architecture and multi-task learning on TITAN)

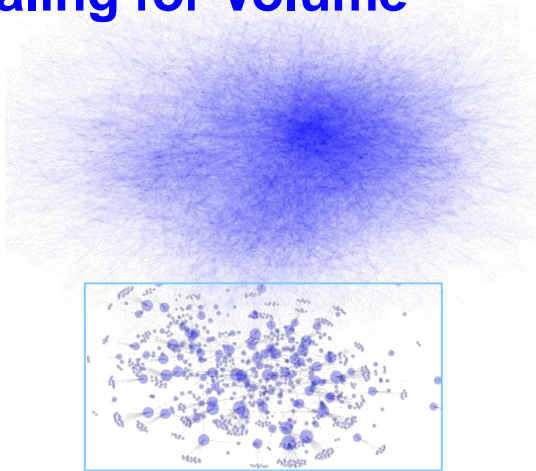
Why EAGLE ?

- Graph analytics and Mining
 - Scale for volume (Billions of nodes and edges),
 - Scale for heterogeneity (Multiple types of nodes and edges)
 - Scale for real-time inference
- Semantic Reasoning with Data and Meta-data
 - Inability limiting potential discoveries
 - Potential with unstructured data
 - Data + Meta-data graphs (ontologies, XML etc.) hard to represent as adjacency matrices.
- Associative-memory based learning
 - N-d-k problem (incremental k)
 - Understand power of in-memory computing

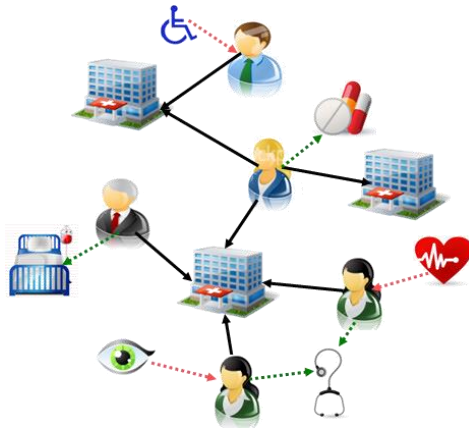
Vision: Offer a graph database and analytical services for mining massive heterogenous graphs

What is EAGLE ?

Scaling for Volume




Scaling for Heterogeneity



EAGLE algorithms are offered as a RESTful service

EAGLE: Eagle 'IsA' Algorithmic Graph Library for Exploratory-Analysis

APOLLO
Universal RDF Integration Knowledge Appliance
Retrieval Module



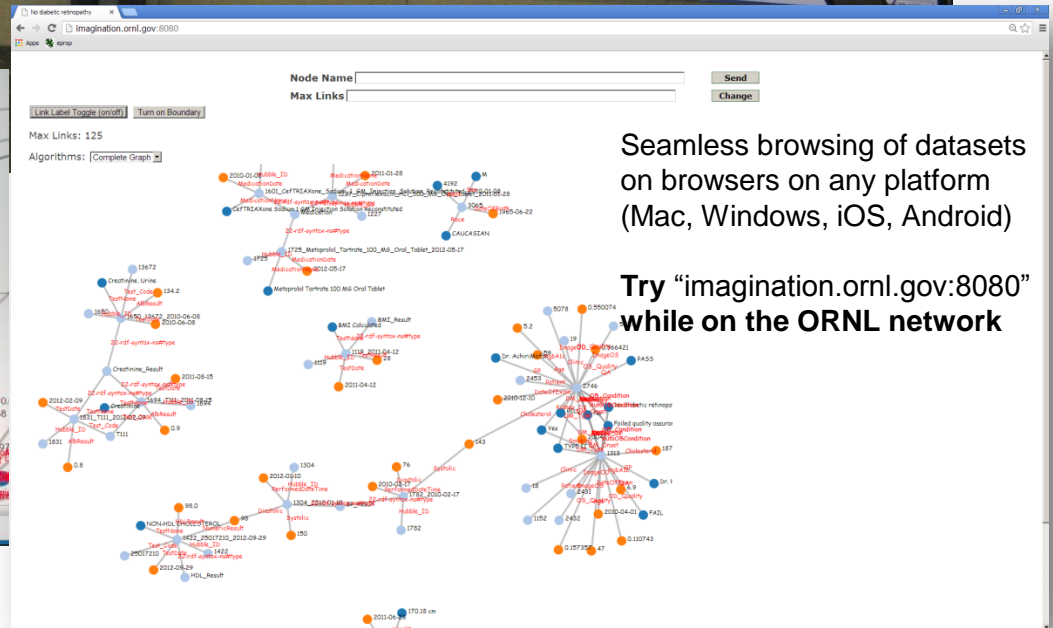
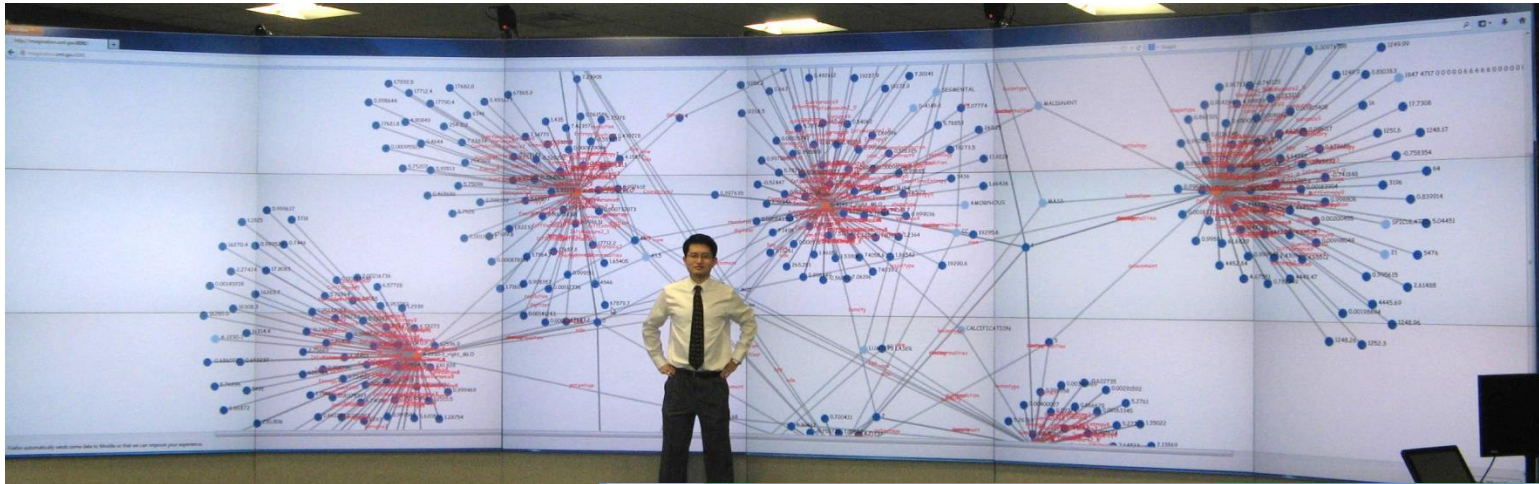
EAGLE
Knowledge Discovery and Graph Mining Library
Exploratory Module

Centrality
Degree, Closeness, Eigen Vector, Betweenness
Flow
PageRank, Belief Diffusion and Propagation
Graph Operations
Ego Graph Extraction, Graph Difference, Fusion Power, Shortest Path
Graph Exploration
Dominant Edges, Dominant Nodes, Histograms by node and edge types
Graph Clustering
Degree-Stratified clustering coefficient, Peer-pressure clustering
Graph Compression/ Filtering
Snowball sampling, Collaborative Filtering, SUBDUE

Relationship Analytics Platform
Enabling Discovery-by-association

Converts “in-memory storage and retrieval” appliance to a reasoning/discovery platform.

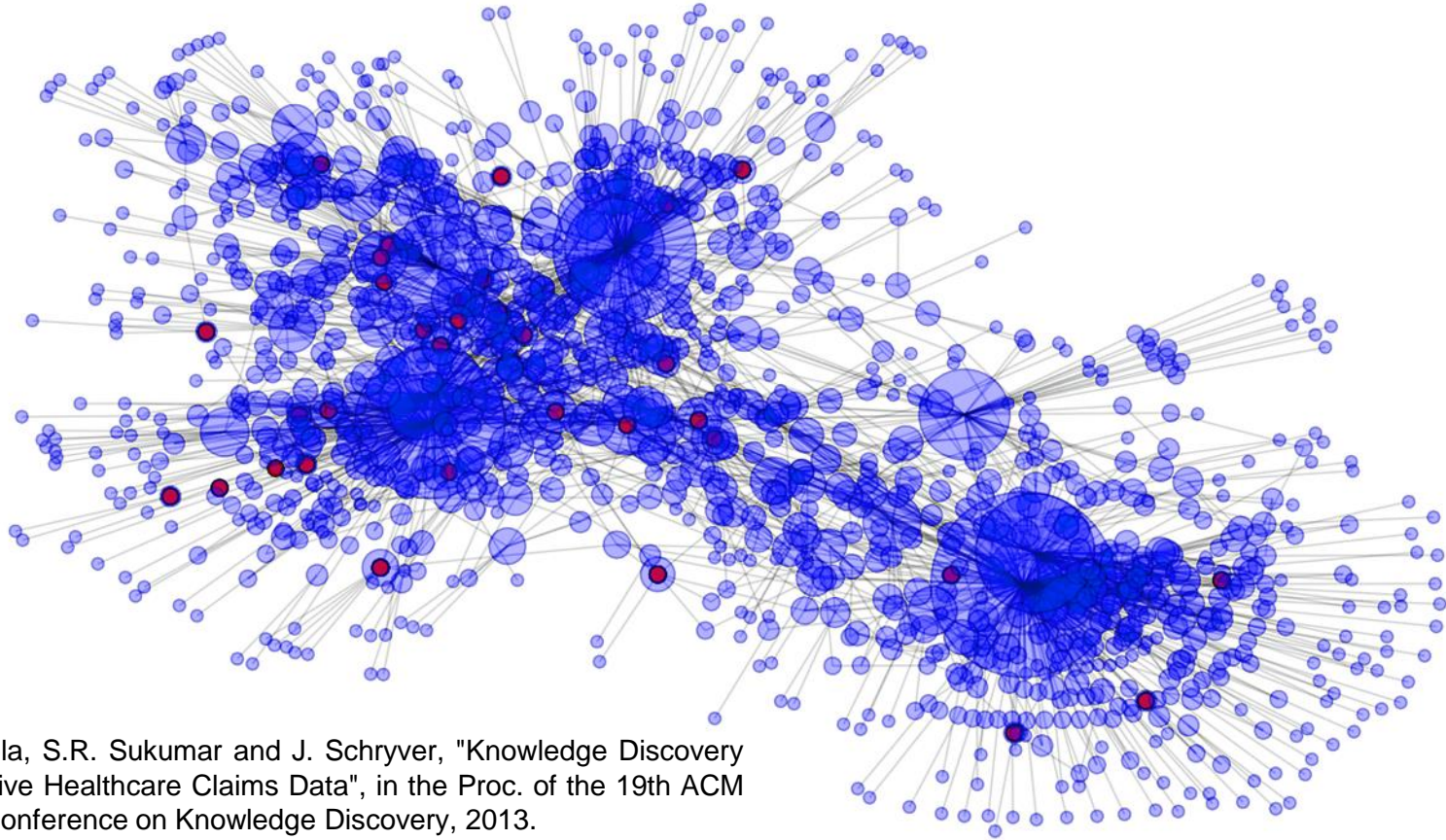
What can EAGLE do for you ?



Seamless browsing of datasets
on browsers on any platform
(Mac, Windows, iOS, Android)

Try "imagination.ornl.gov:8080"
while on the ORNL network

Use Case: Pattern Recognition



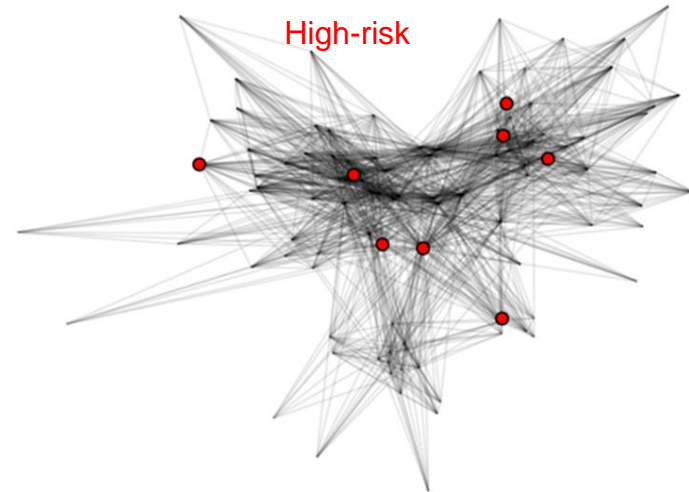
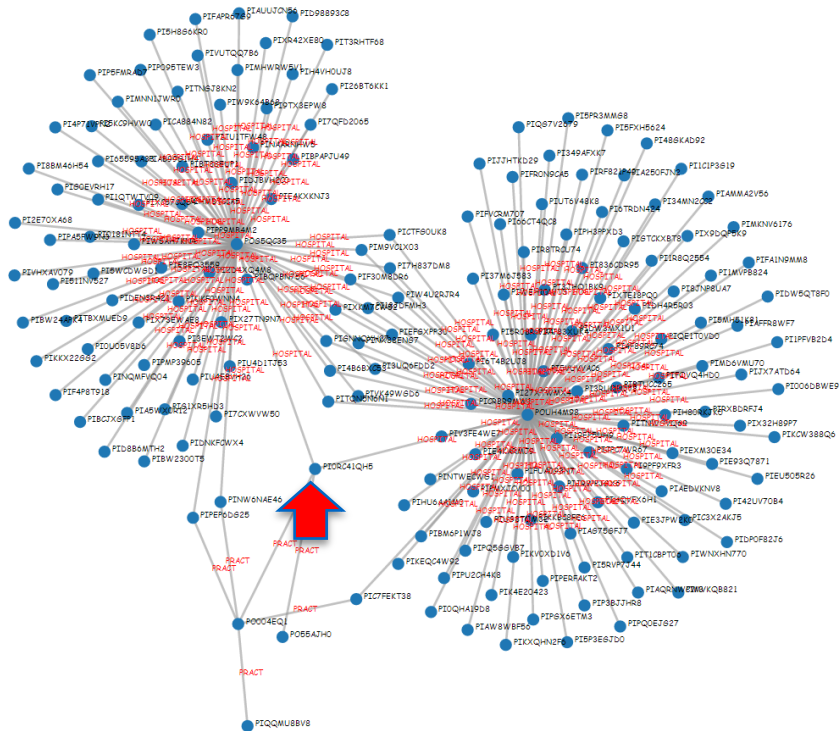
V. Chandola, S.R. Sukumar and J. Schryver, "Knowledge Discovery from Massive Healthcare Claims Data", in the Proc. of the 19th ACM SIGKDD Conference on Knowledge Discovery, 2013.

- Given a few examples of fraud (important activity) , can we
- (i) Discover patterns typically associated with suspicious activity?
 - (ii) Extrapolate such high-risk patterns for investigation and fraud prevention?

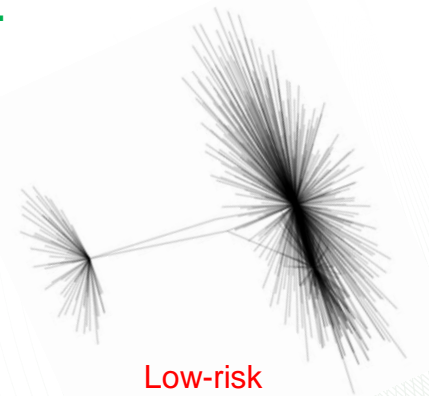
Use Case: Pattern Recognition (contd.)

Insight from Patterns

Affiliations to multiple hospitals, owning private and group practice are strong indicators of potential suspicious activity.

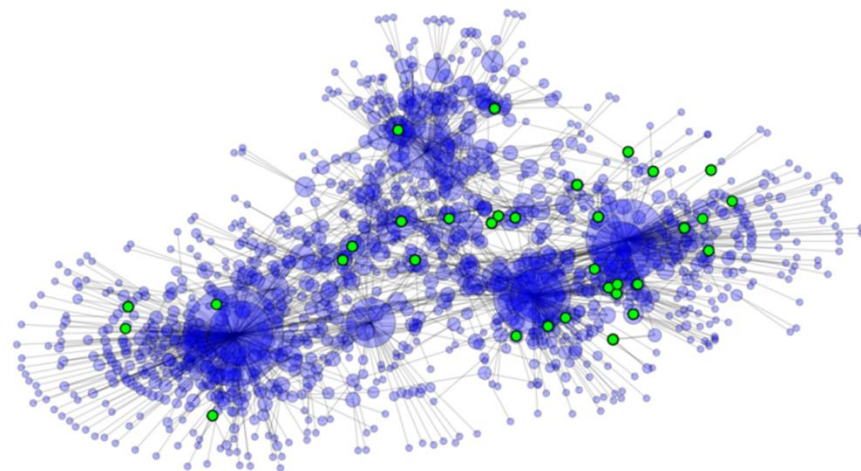
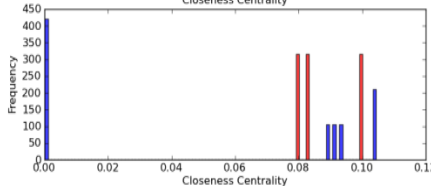
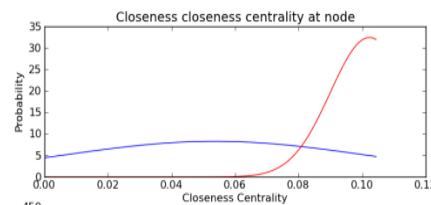
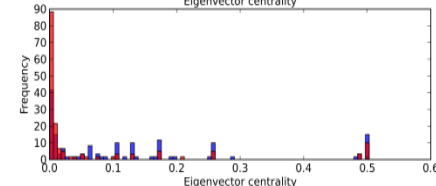
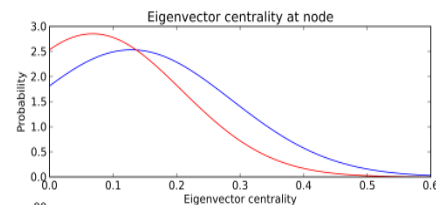
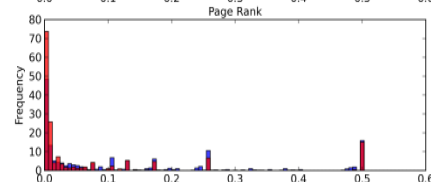
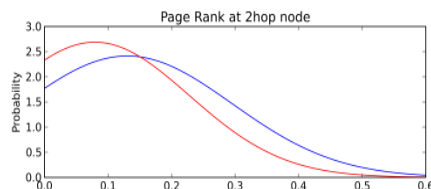
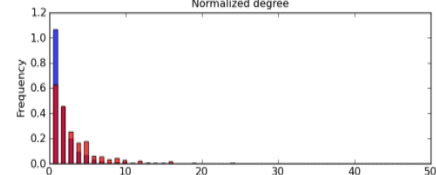
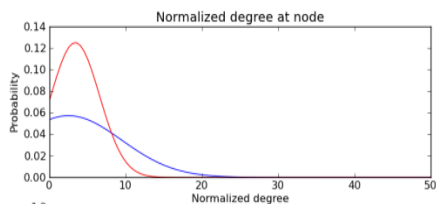


High risk “referral” patterns are Y networks and triangles (above), low risk patterns are star-shaped (below).

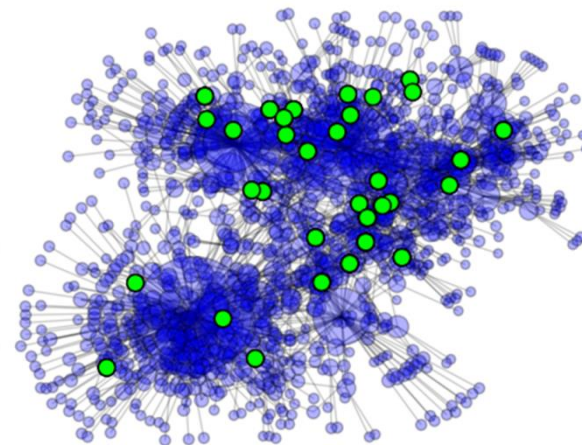


Use Case: Pattern Recognition (contd.)

Understanding relationship patterns
using mathematical models

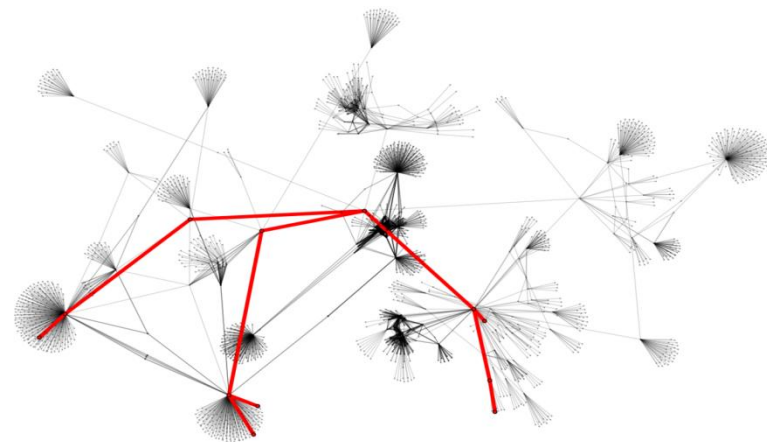
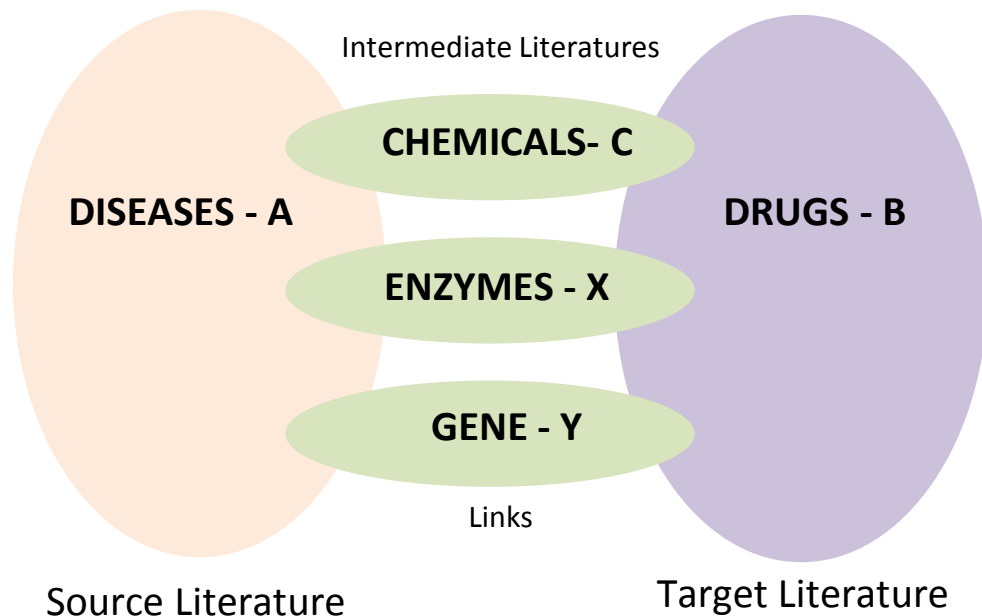


Extrapolating for new
hypothesis/investigation



Use Case: Literature-based Discovery

On going pilot with the National Library of Medicine



Meta-data links after disparate data integration. The red-lines indicates a non-obvious data-element link.

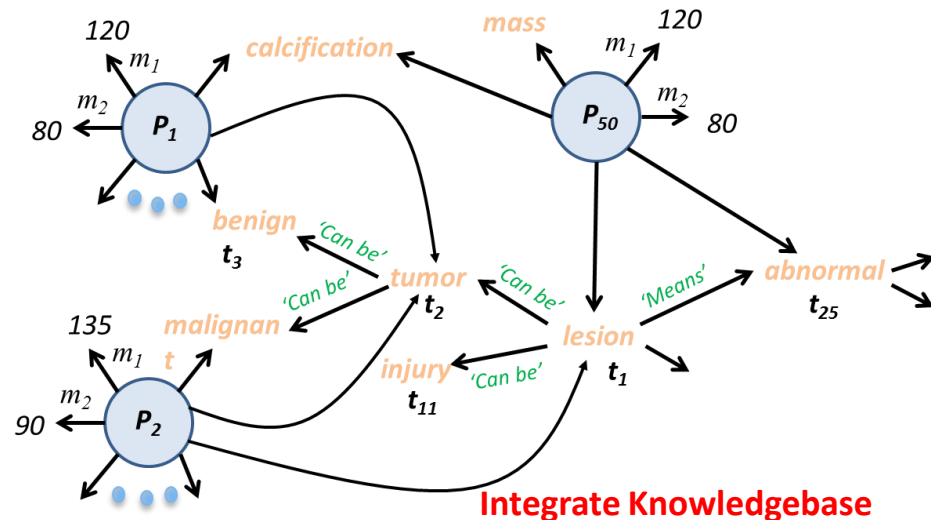
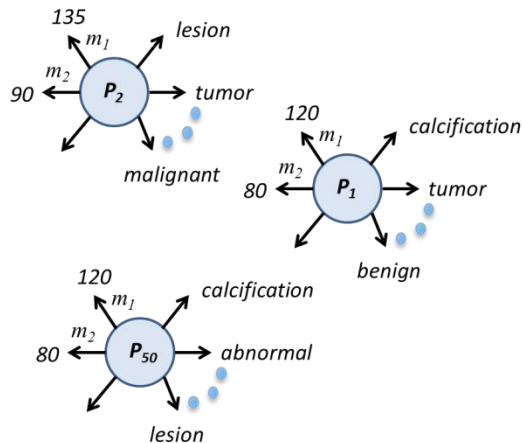
S. R. Sukumar and R. K. Ferrell, "Collaboration When Working With Big Data: Recording and Sharing Analytical Knowledge Within And Across Enterprise Data Sources", Information Services and Use Journal, 2013.

Given data and meta-data as knowledgebase from different domains, can we

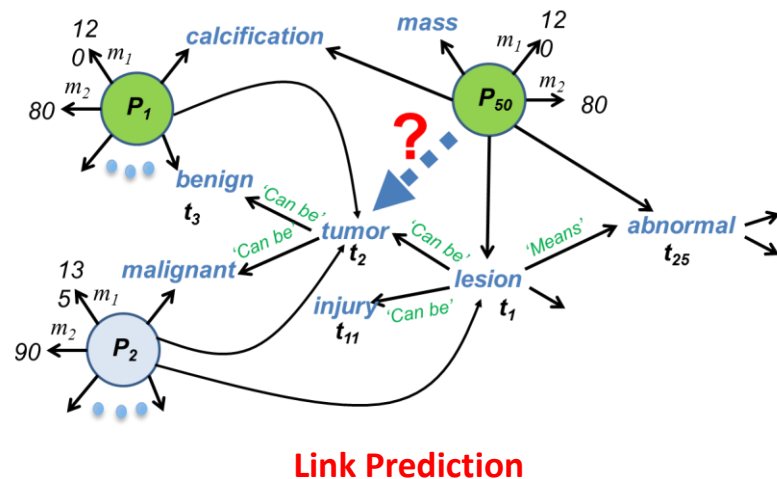
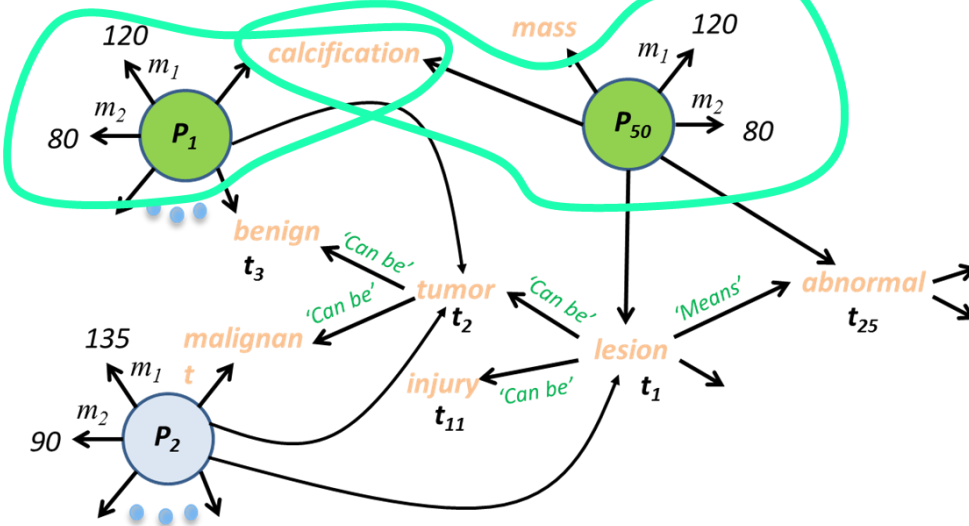
- (i) Discover new relationships of entities between domains ?
- (ii) Automatically extract and prioritize discovered relationships for clinical or subject matter expert validation?

Use Cases: Multi-structured Data Analysis

Convert data into RDF



Finding Similar patients



S. R. Sukumar, and K. C. Ainsworth. "Pattern search in multi-structure data: a framework for the next-generation evidence-based medicine." In *SPIE Medical Imaging*, pp. 903900-903900, February 2014.

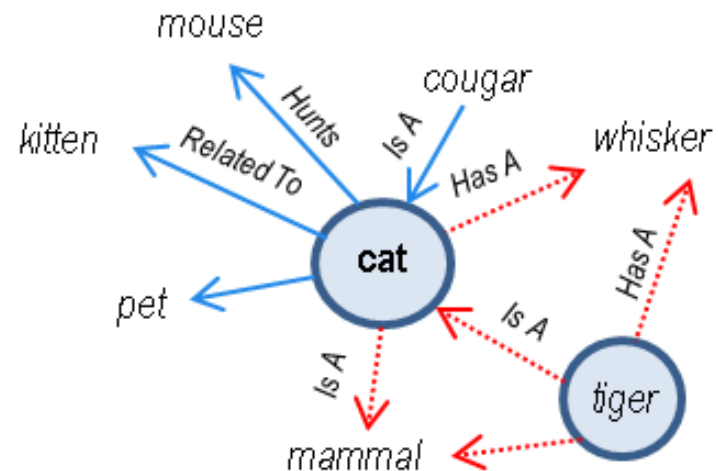
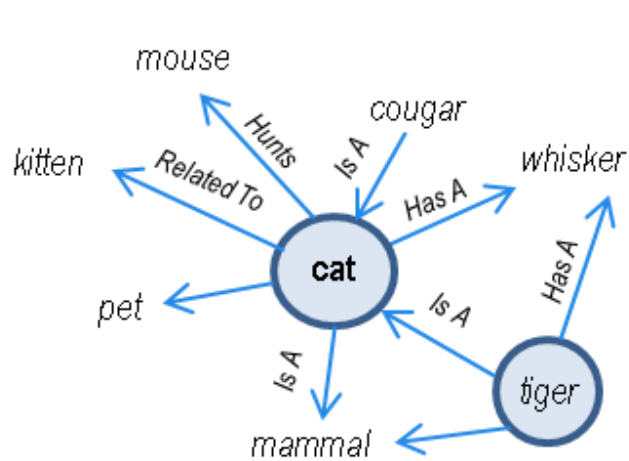
Use Cases: Semantic Pattern Analysis

A verbal-fluency exam:

Neuro-psychiatrist: Name as many animals as you can in 60 seconds ?

PTSD Patient : racoon, lion, elk,.....

Normal Patient: cat, tiger, dog, horse,....



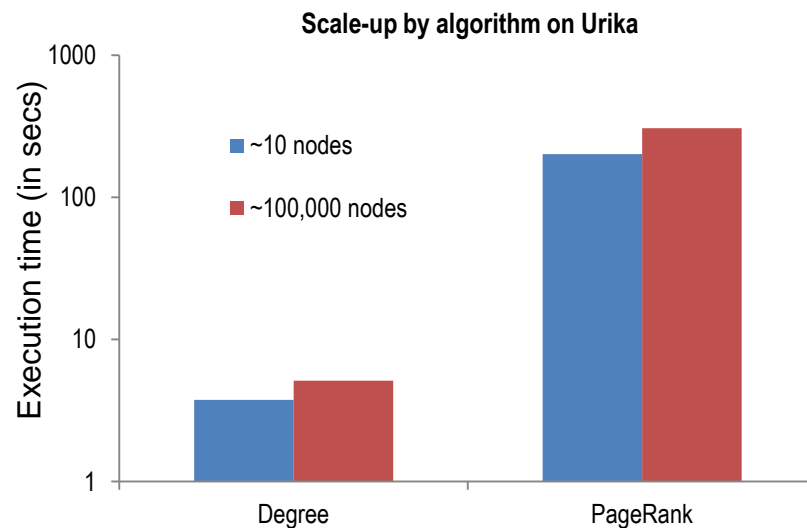
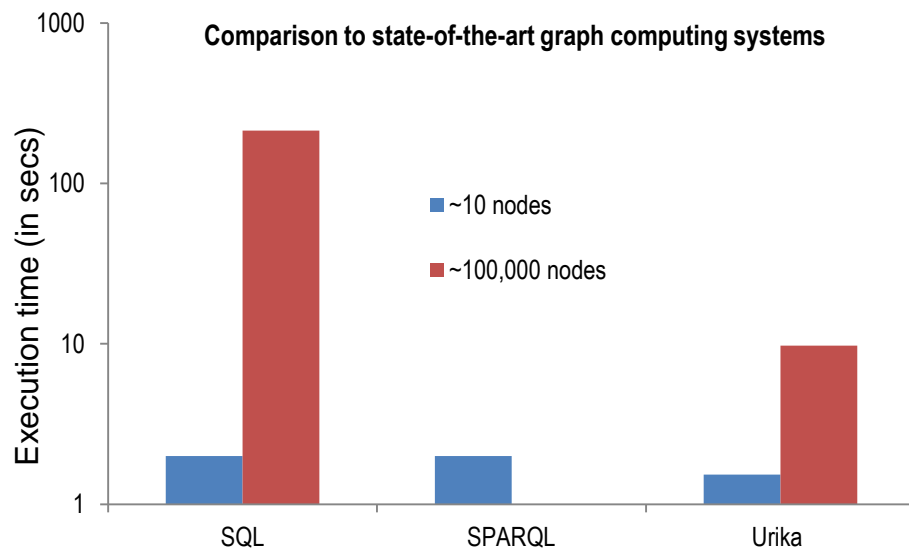
Can we host all of human common sense in-memory (of a computer) to evaluate patterns of thought progression during verbal fluency exams ?

Sukumar, Sreenivas R., Keela C. Ainsworth, Tyler C. Brown, "Semantic Pattern Analysis for Verbal Fluency Based Assessment of Neurological Disorders", IEEE 4th ORNL Biomedical Science and Engineering Conference, May 2014.

Key Capability

- Data-driven approach to automatic pattern search and discovery on data represented as massive heterogeneous graphs.
- Generic EAGLE Toolbox presents algorithms as APIs with visualization capability – extendable to applications beyond healthcare and intelligence.
- Scalable graph computing:
 - **10-100X on graph-theoretic algorithms (comparable hardware).**
 - **First of its kind reasoning with semantic data.**

Scaling for Volume (Preliminary benchmarking results)



Three Analytical-Support Toolkits

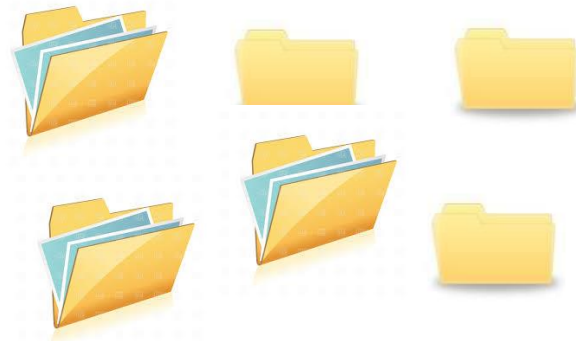
- Still in R&D stage (Use-cases and ideas very welcome)
 - EAGLE : The exploratory discovery module on ORNL's Apollo (YarcData Urika)
 - SpotHadoop: Hybridizing HPC and Big Data community (Tested on: Smoky, Rhea, EOS Troubleshooting: TITAN)
 - iLearn: Machine Learning using heterogeneous computing (Implementing self-learning methods such as deep-and multi-task learning on TITAN)

Why SpotHadoop ?

- Users are moving large datasets around for analysis purposes
 - Frustrated that data is too big to use in their favorite tool
 - Data takes too long to be migrated for analysis.
 - Have to pay \$\$\$\$\$s to services like Amazon
 - Moving data – often translates to – provenance leak and publication tracking issues.
- Big Data Developer Community >>> HPC Developer Community
 - Tremendous progress in analytical tool development with Big Data stacks



I have 1 PTB file on LUSTRE !



I have 500 1 TB files !

What is the histogram of average temperature ?

Hadoop on HPC: Good idea ?

➤ Hadoop opens suite of analytical tools

- Structured and Unstructured
 - Hive, Pig, Spark (SQL, MapReduce)
 - HBase, Cassandra (No-SQL)
 - Mahout, Pegasus (Machine learning)

We don't have to develop or re-implement algorithms and data analytics tools in MPI.

➤ Storage cost in cloud vs. HPC

- Cost of 250 PB for 1 year
 - Amazon AWS : Cray HPC : : 30 : 11

We don't want to pay twice as much!

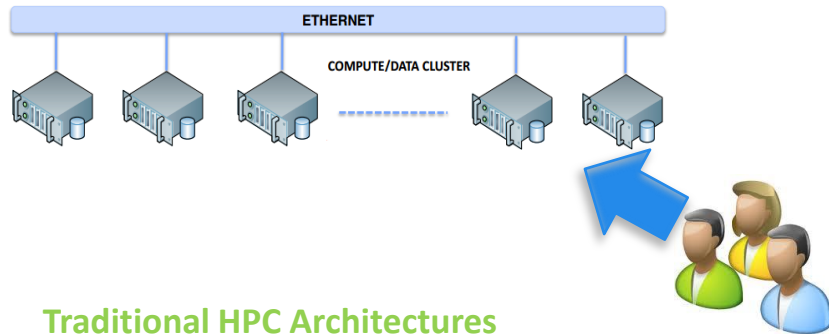
➤ Everyone is doing it !

- NERSC: MARIANNE
- SDSC: myHadoop
- UPSC: STEM
- ORNL : ?

We don't have to move data outside HPC storage system.

Hadoop on HPC: Bad Idea ?

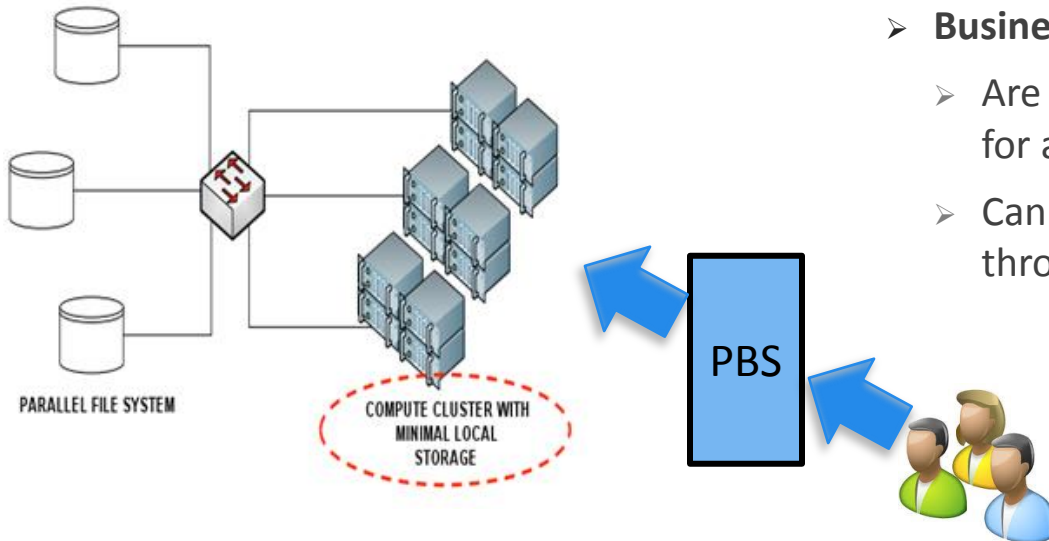
Initial Philosophy behind Hadoop



What are the challenges ?

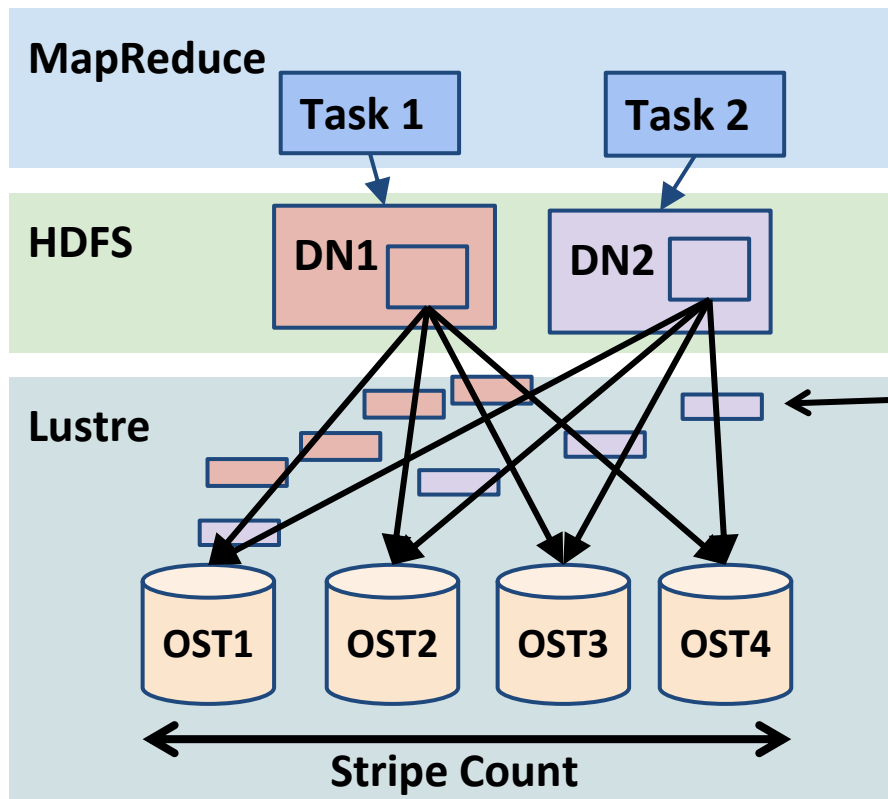
- **Performance : Architectural differences**
 - Hadoop expects data from local storage
 - Is Spider's bandwidth good enough ?
 - How to utilize Cray's Gemini capability (CCI) ?
- **Operational challenges**
 - **Business model**
 - Are we willing to dedicate HPC resources for a Hadoop cluster ?
 - Can we dynamically configure Hadoop through PBS job submission system ?

Traditional HPC Architectures

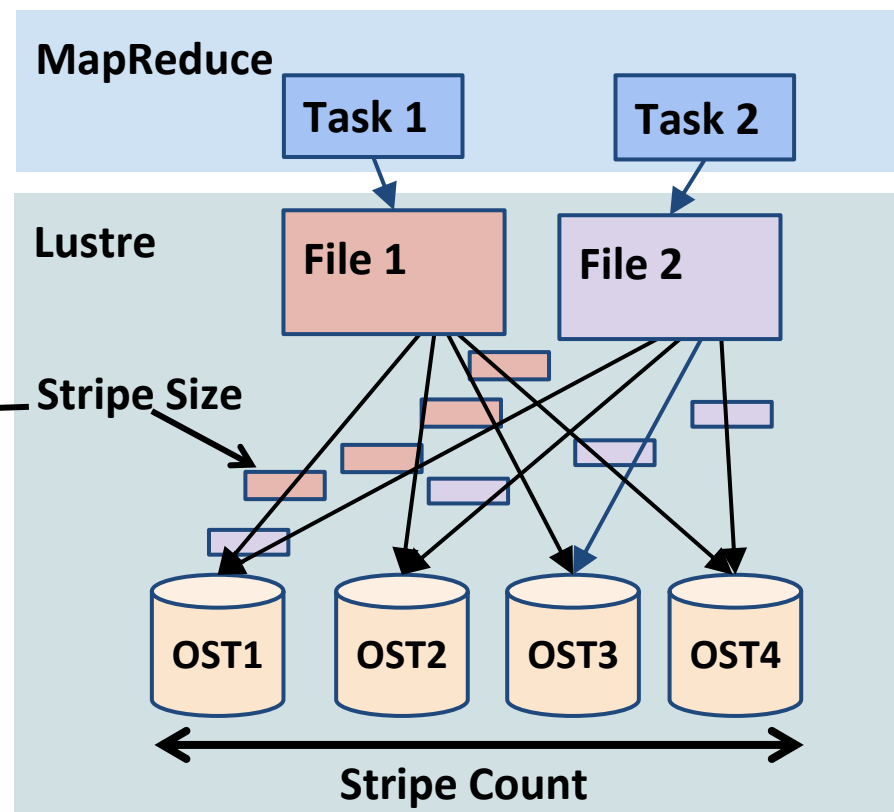


How did we make it work ?

- Hadoop using HDFS on LUSTRE
 - Additional checksum calculations affect performance
 - Two layers of meta-data service



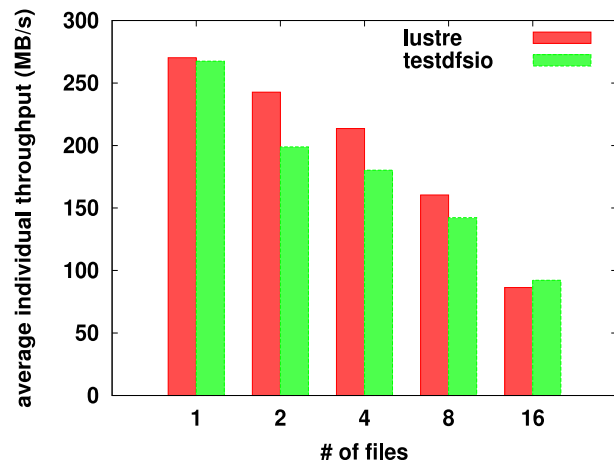
- Hadoop over LUSTRE
 - *Nearly 100 times faster*



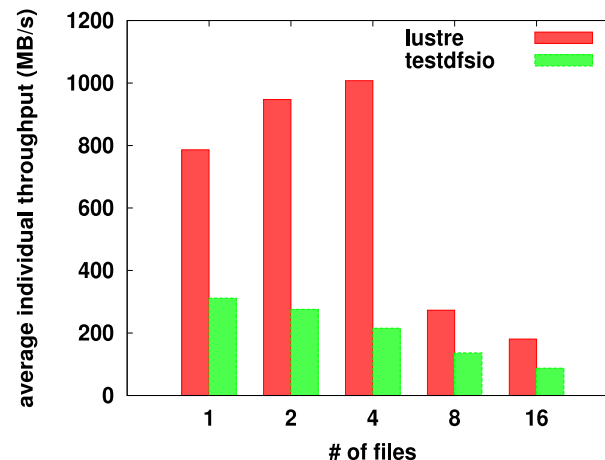
Was SPIDER good enough ?

SPIDER is good enough.

Experiment: Performance (accessing n files from one node)



(a) Write throughput



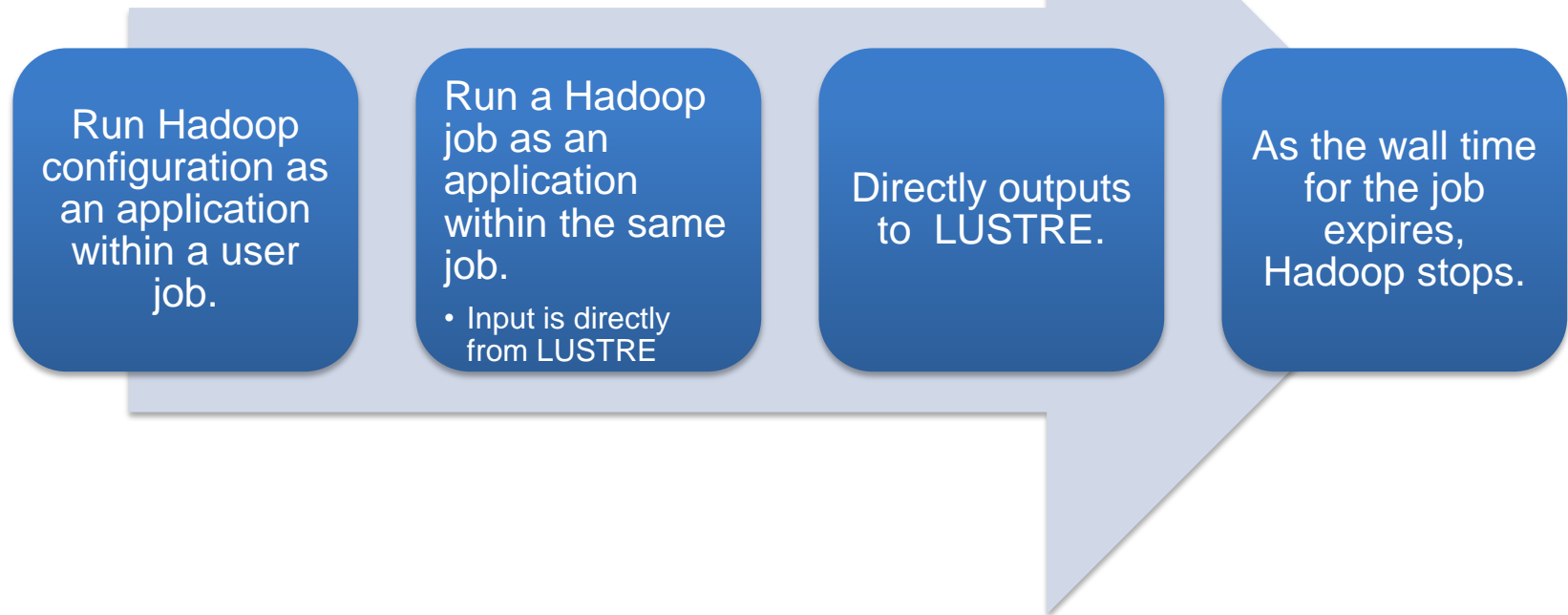
(b) Read throughput

Less than 18% of deviation from raw lustre.

One file cases represent the overhead from additional software stacks.

What is SpotHadoop ?

- On-demand Hadoop



• Where are we on testing SpotHadoop ?

- Smoky: Yes Rhea: Yes EOS: Ongoing Titan: Ongoing (CCI Issues)

Is SpotHadoop better than AWS ?

Preliminary Benchmarking Results

- Workloads
 - TestDFSIO : basic I/O performance benchmark for Hadoop jobs.
 - 1, 4, 8, 16, 32 map tasks and each map task read/write a 1GB file.
 - Terasort: an example Hadoop job with both map tasks and reduce tasks.

System Specification: Comparable Hardware

Rhea

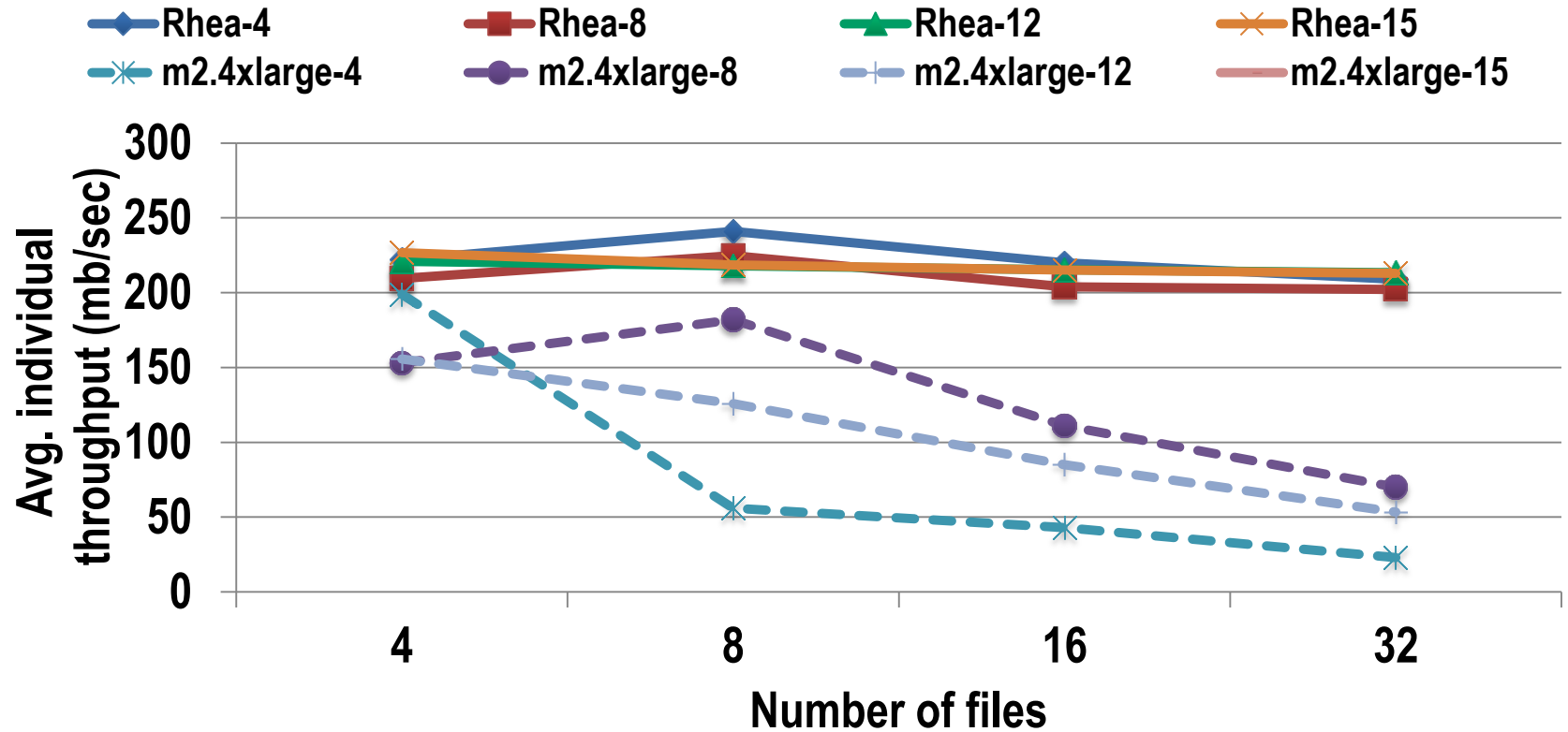
- Processor: 16 CPU cores with hyperthreading (32CPU at OS)
- Memory: 64GB RAM
- Network: Ethernet and Infiniband
 - We used Ethernet interface
- Storage:
 - Center-wide Lustre-based file system.
 - No local storage.
 - Lustre filesystem

M2.4xlarge in Amazon Web Service

- Processor: 8 virtual CPUs
- Memory: 68.4GB
- Network : Ethernet
- Storage:
 - Hadoop data : local storage (840GB x 2 hard disks)
 - Root: networked block device service
 - Hadoop Distributed File System (HDFS)
- Price: \$0.980 per instance hour

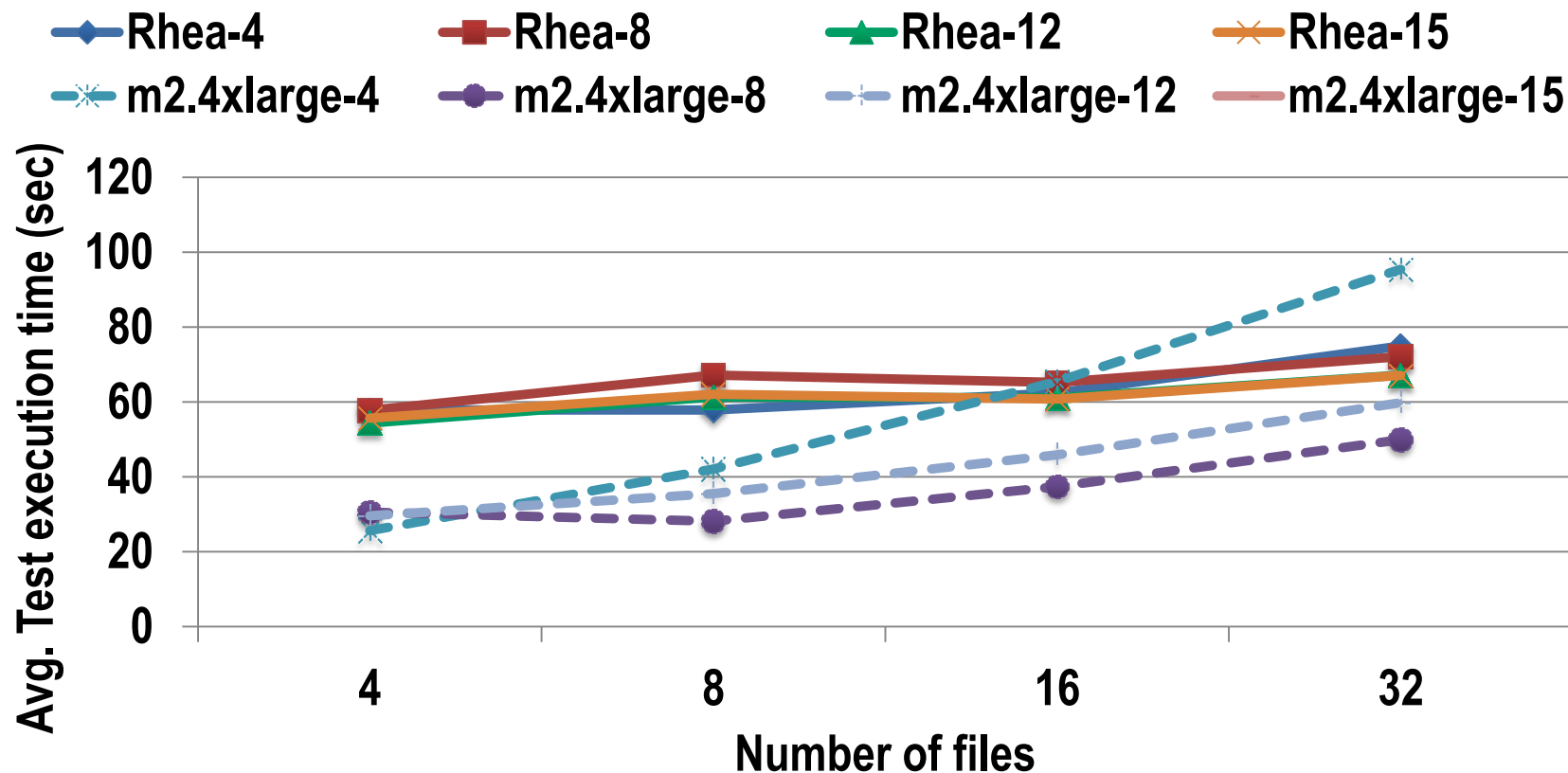
Evaluation (read throughput)

TestDFSIO Read 1GB files



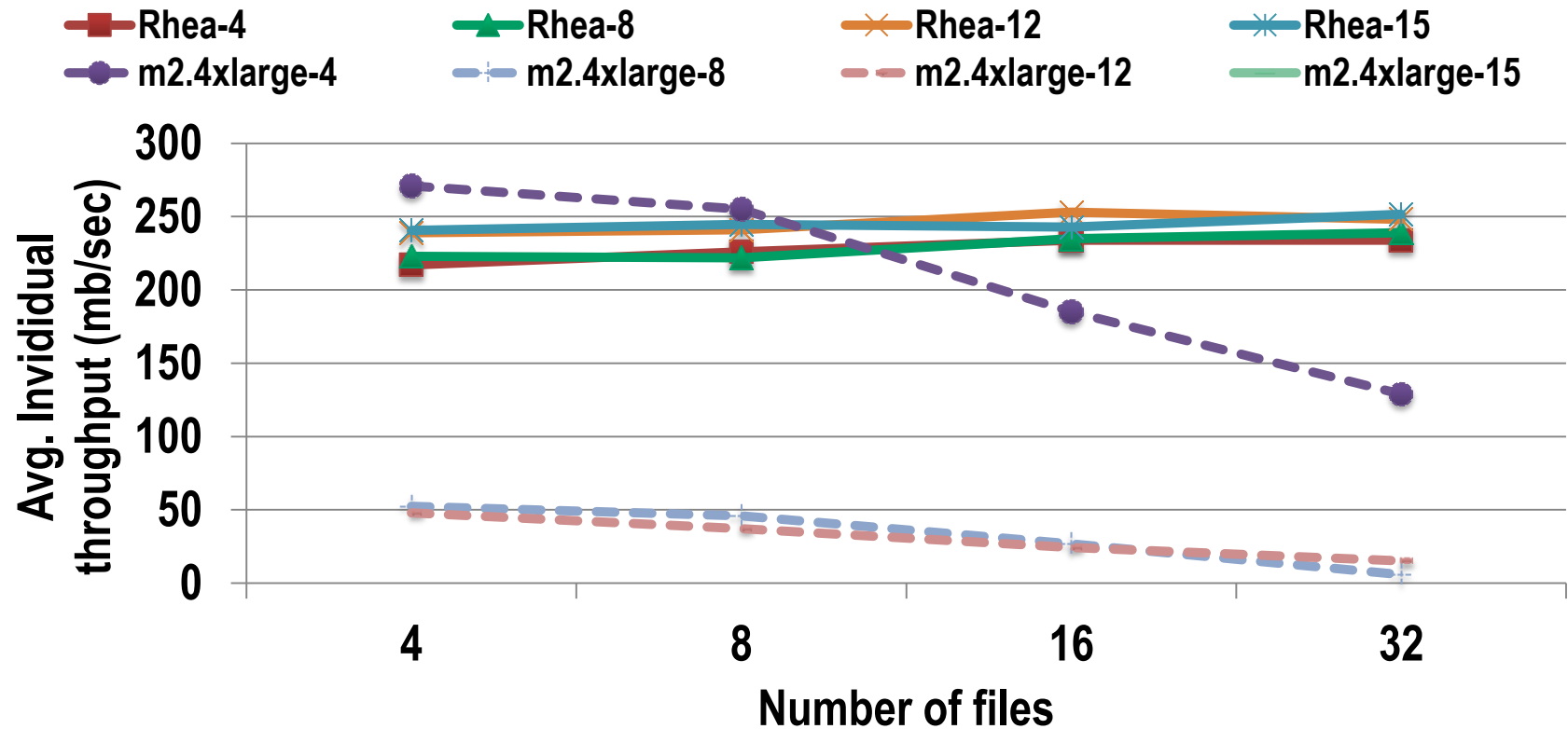
Evaluation (read running time)

TestDFSIO Read 1GB files



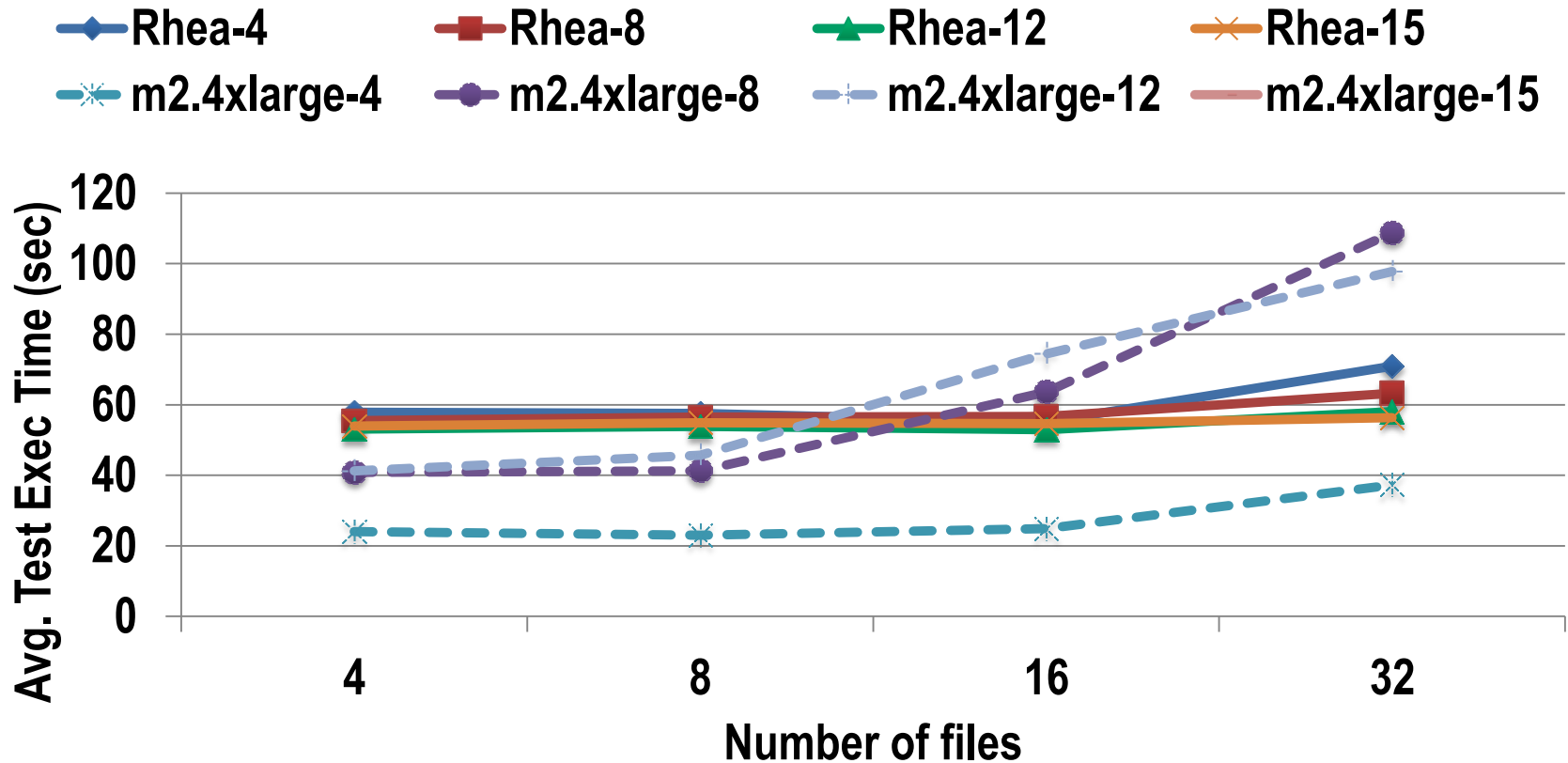
Evaluation (write throughput)

TestDFSIO Write 1GB Files



Evaluation (write running time)

TestDFSIO Write 1GB Files



Ongoing Work: Big Data on Big Iron Benchmark

Looking for collaborators across facilities

Architectures:

- In core-processing
 - TITAN (Cores)
 - TITAN (Cores + GPU)
- In storage processing
 - Hadoop + Mahout (AWS)
 - Greenplum + Madlib
 - Hadoop on Rhea, EOS
- In-memory processing
 - Hadoop + Impala + Pegasus
 - Urika (SPARQL)
 - Cloud + Jena (AWS and CADES)

Algorithms:

- Retrieval
 - Fetch a random record
 - Get a record with at least 5 joins
- Simple analysis
 - Collaborative filtering
 - K-NN
 - Degree
 - K-means
- Complex algorithms
 - Page-Rank
 - Matrix Inversion (Eigen-value)

Towards designing the future knowledge nurturing discovery architectures....

Thank You

- Questions ?